DOI:
10.1111/ijal.12174

Document Version
Peer reviewed version

Link to publication record in King's Research Portal

Citation for published version (APA):
Tytus, A. E., & Rundblad, G. (2017). Cross-linguistic Semantic Transfer in Bilingual Chinese-English Speakers. *International Journal of Applied Linguistics*, *27*(3), 697-711. https://doi.org/10.1111/ijal.12174

Cross-linguistic Semantic Transfer in Bilingual Chinese-English Speakers

Agnieszka Ewa Tytus[1], Gabriella Rundblad[2]

[1] Corresponding author: King's College London, DEPS, Franklin-Wilkins Building, WBW, Waterloo Road, London, SE1 9NH, UK. agnieszka.tytus@kcl.ac.uk

[2] King's College London, DEPS, Franklin-Wilkins Building, WBW, Waterloo Road, London, SE1 9NH, UK. gabriella.rundblad@kcl.ac.uk

Cross-linguistic Semantic Transfer in Bilingual Chinese-English Speakers

Abstract

The semantic structure of animal and emotion terms was here under investigation. In Study one, Chinese-English bilinguals and monolingual English and Chinese speakers provided similarity judgments. Those, once analysed with correspondence analysis, provided a multidimensional representation of the semantic structure of animal terms; with the greatest level of similarity noted between two bilingual structures. In Study two, a within subjects design was employed and the participants evaluated levels of similarity between pairs of emotion terms. The greatest level of similarity was recorded between the bilingual judgments provided in English and the monolingual English ones. The aggregated findings have demonstrated that the bilingual semantic structures are dynamic, possibly due to the constant interaction between two languages, which in consequence may lead to creation of *semantic interlanguage*.

Keywords
bilingualism, semantic domain of animals, semantic domain of emotions, semantic judgment task, semantic interlanguage

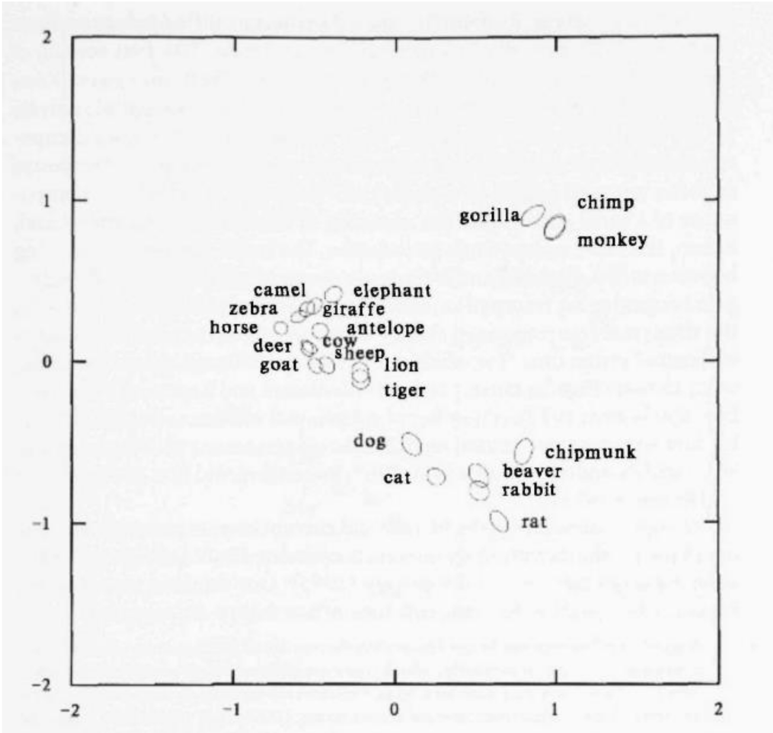
Introduction

'The bilingual is not two monolinguals in one person'. This influential observation is part of the title of a research paper written by Francois Grosjean over 25 years ago (1989). It denotes that a bilingual person should not be considered as simply comprising two monolingual counterparts, but rather, should be treated as a 'unique and specific speaker-hearer' (Grosjean, 1989:3). The uniqueness of a bilingual individual can, for instance, be related to the fact that they draw on the knowledge of two or more linguistic repertoires and create a new, often in-between representation that does not fully resemble either of the monolingual systems; in short, they develop an *interlanguage* (Selinker, 1972). This in-between representation is illustrative of the dynamic and idiosyncratic nature of language learning and it can demonstrate itself at the level of phonology (e.g., Dickerson, 1975), syntax and morphology (e.g., Gass, 1984), or pragmatics (e.g., Kasper and Schmidt, 1996). The extensive research available on interlanguage, mostly conducted in the 70s and 80s, does not, however, extend to direct investigations of the semantic level of representation. This might be related to the fact that the object of study of semantics is rather elusive, especially when compared to other realms of the linguistic inquiry. More recent studies conducted by Ameel et al. (2005; 2009) are among a few that focused specifically on this level of representation in bilingual populations. These groups of researchers investigated the domain of drinking vessels and household dishes and showed that bilinguals undergo a process of semantic convergence. That is, the bilingual category centres are located closer to each other compared to the monolingual ones and they are located in-between the monolingual category centres on a multidimensional scale (MDS) map.

This shortage of research is rather surprising as an individual acquiring a second language does not simply have to learn a new set of labels that can be matched to already existing concepts, but as indicated by Pavlenko (2009) L2 learning leads to a conceptual restructuring. Furthermore, if we consider bilinguals who are proficient in their two languages rather than language learners, we need to account for the constant interaction between two languages. The impact of first language (L1) on L2 is well-documented, e.g., Swan (1997) vocabulary acquisition and use, Harrison and Krol (2007) phonological processing, or MacWhinney (1997) grammatical competence. Also, L2 influence on L1 has been well-evidenced (see Cook, 2003 for a review). Since both languages are active at all times and the one which is not in use is inhibited (Green, 1998), we can assume different interactional patterns, both temporary as well as long-lasting, leading to changes at different levels of language representation, also at the semantic level. That is, since bilingual speakers have to accommodate language specific lexicalised concepts (mappings of forms to conceptual meanings); they might create a form of a *semantic interlanguage*. For example, a Chinese-English speaker has to manage quite distinct lexicalised concepts when it comes to, e.g., (1) a dining table (in China, normally a round piece of furniture with a rotating inner part for serving food compared to a typical English square table), (2) connotations of the colour *red* (in China this is associated with good luck, prosperity, good fortune compared to English when it refers to alert, danger or passion), or (3) the concept of waiting for someone outside (in China one normally squats while waiting compared to standing or sitting down/perching on something). It has been demonstrated that language/culture specific concepts are accessed quicker from the congruent language than the incongruent one (Jared, Poh, and Paivio, 2013), but little is known about the interaction between the language specific lexicalised concepts. That is, it is not certain if a Chinese speaker of English restructures the concept of ‘squatting as waiting’ in favor of ‘perching as waiting’ with high levels of English language proficiency and extended exposure to English language and culture. The importance of investigating this level of representation and the scarcity of available findings led us to explore the process of semantic change in two domains: animals and emotions with the use of a semantic judgment task as it allows for the investigation of the *structure of semantic domains* (Romney, Moore, and Rusch, 1997). Furthermore, we focused on an under researched group of speakers, i.e. a Chinese-English bilinguals by following both a between subjects design (Study one) and a within subjects design (Study two).

Animal Scaling Studies

A *semantic domain* can be understood as an organised set of words that refers to a single conceptual category, such as kinship terms, colour terms, or emotions terms; whilst, the *structure of a semantic domain* may be described as the arrangement of the co-hyponyms relative to each other represented in Euclidean space (Romney et al., 1997). The structure of a semantic domain is derived from a judged-similarity task, which can take several different forms. For instance, the participants are presented with pairs of words and are requested to indicate on a scale how similar or dissimilar the pairs are. Alternatively, they are shown triplets of words and are asked to point to the word that is least similar to the other two (a form of an Odd One Out task). Also, some studies (e.g., Holmes and Wolff, 2013) relied on sorting stimuli into piles/distinct categories. Once the stimuli of a given semantic domain are ranked in terms of similarity, a multidimensional scaling (MDS) analysis allows for the production of a spatial representation of the semantic relationship between the terms (Herrmann and Raybeck, 1981). An example of a MDS solution taken from the domain of animals is presented in Figure 1. The map can be interpreted by accepting that the terms that are judged by the participants as more similar are positioned closer to each other than those seen as less similar (Romney et al., 1997). For instance, it can be seen on the map below that animals, such as gorilla, chimp, and monkey were judged as more similar to each other, compared to, say, horse and rat, as these are much further apart.



The previous animal scaling studies, i.e., Romney et al. (1995) or Herrmann and Raybeck (1981), were conducted with monolingual speakers. Romney et al. (1995) worked with a group of English speakers and presented a model of the semantic structure of the domain of animals, which, according to the researchers, could reflect the cognitive representation that the individuals bear in their minds. The study conducted by Herrmann and Raybeck (1981) added a further dimension to the investigation of the animal terms by making a comparison of the similarities in meaning between six cultures, i.e., Spanish, Vietnamese, Chinese, Haitian, Greek, and American. The researchers noted that since the position of animal terms did not correspond on the MDS maps across the six investigated cultures, this might be reflective of salient cultural differences (Herrmann and Raybeck, 1981:203). Drawing on the above findings, i.e., on the fact that the judgments provided by speakers of different cultures may reflect culture-specific meaning, we have assumed that bilinguals who speak two languages on a daily basis may categorize animal terms, though only to a small extent, differently in their two languages. Also, we have expected the bilingual categorization patterns may differ from those offered by monolingual speakers of each respective language, especially is very distinct languages such as Chinese and English are considered.

Emotion Scaling Studies

So as to be able to generalise across other semantic domains, for the second Study we selected the semantic domain of emotions. In comparison to the domain of animals, these do not have concrete referents, they cannot be matched neatly across language and they have cultural boundaries (Wierzbicka, 1992:287). The available emotion scaling studies, e.g., Alvarado and Jameson (2011), Moore et al. (1999) and Russell (1983) suggest substantial commonalities between the emotion terms between different languages. For instance, Moore et al. (1999), who worked with Chinese, Japanese and English speakers, reported that there is strong similarity across the emotion lexicons. That is, the participants in the three languages assigned similar meanings to the investigated terms and hence, the finding points to the universality of emotion across cultures. The same pattern of results, i.e., a shared understanding of emotion terms, was shown in a study conducted by Alvarado and Jameson (2011:974) with groups of monolingual English and Vietnamese speakers as well as bilingual Vietnamese-English ones. As noted by the researchers, some 'culture-specific differences were found where they were expected'. For instance, a difference in the conceptualisation of *anxiety* was observed, i.e., on the monolingual English MDS map and on the map pertaining to bilinguals responding in English, *anxiety* was clustered closely with *excitement*, *love* and *happiness*. This distribution was, however, not present on the other maps, i.e., on the monolingual Vietnamese and the bilingual Vietnamese one. Alvarado and Jameson (2011) speculated that the difference might have occurred due to the fact that English speakers differentiate between emotions that are discrete, such as: fear, anger, or happiness, and states that are reflective of mood, e.g., calmness, depression, or anxiety, or arousal states such as relaxed, tense, or nervous. Finally, the influential study conducted by Russel (1983) with Gujarati, Croatian, Japanese, Chinese, and English speakers demonstrated several pancultural properties, two of which being (1) a circular order of the emotion terms in (2) a two dimensional space, with the dimension referring to: pleasure-displeasure and arousal-sleep. The bipolar pleasure-displeasure dimension was shown in a number of early cross-linguistic studies (e.g. Yoshida et al., 1970; or Fillenbaum and Rapoport, 1971). Other elements, however, including the distribution of terms on the maps differed. This might be related to a difference when displaying and experiencing emotions. For example, Tsai, Simeonova and Watanabe (2004) suggest that Chinese and Americans describe emotional experiences differently, i.e. Chinese speakers tend to use more somatic, e.g. *dizzy* and social words, e.g. *friend* compared to Americans. Taken the above described differences into account, we set off to investigate the semantic structure of emotions in bilingual Chinese-English speakers.

Overview of the Studies

To investigate the processes of semantic change, two studies with Mandarin Chinese-English speakers and English and Chinese monolinguals were conducted exploring the semantic structure of animal and emotion terms. Study one focused on the semantic domain of animals and aimed to compare similarity judgments in the form of multidimensional scaling maps. This semantic domain was chosen due to the fact that the exemplars are concrete entities that have well defined physical characteristics, such as: size, shape, and colour. Furthermore, "animals were also always present in the environment in which humans evolved so that the evolution of visual mechanisms for their detection and characterization can be assumed" (Romney and Moore, 1998:316). If the process of semantic change manifested itself in this group of bilinguals, we can hypothesised that bilingual judgments provided in Chinese should be most similar to those offered by bilingual participants in English, due to the process of constant interaction between the two systems. Also, we can infer that this interaction will make the judgments differ from those offered by monolingual English and monolingual Chinese groups.

In Study two, the purpose was to extend the findings to another semantic domain, i.e., the domain of emotions. By modifying the research design to both between subjects as well as a within subjects one, we aimed at testing the same two hypotheses stated above. Also, based on the theoretical proposal made by Gentner and Boroditsky (2001) regarding the fact that cross-linguistic variation should be greater for more abstract domains than for more concrete domains labeled by common nouns we derived specific experimental predictions. That is, we have expected a slightly greater degree of similarity in the domain of animals and more intercultural variation in the domain of emotions. The overall aim of the two Studies was to demonstrate that the interaction between the two languages may lead to gradual, small however noticeable, changes at the semantic level of representation and as a consequence it may lead to a creation of a semantic interlanguage.

Study one

Method

Participants

Forty Mandarin Chinese-English speakers took part in the task, half of whom provided judgments in English, whereas the other half gave their answers in Chinese. All the participants were from Mainland China and were enrolled on undergraduate (n = 34) and postgraduate courses (n = 6) at the University of Hong Kong. They were between 18 to 25 years old. Their English language proficiency was evaluated with the use of a self-rating questionnaire (Table 1). No statistically significant difference was reported for the levels of English proficiency between the two groups, $M_{levelCh}$ = 3.08 (.33) vs. M_{levelE} = 2.95 (.49), t (38) = 1.040, p > .05. Also, no difference emerged with regard to the age of English language acquisition, i.e., M_{ageCh} = 8.45 (2.13) vs. M_{ageE} = 7.85 (2.49), t (38) = .816, p > .05.

	listening	**speaking**	**reading**	**writing**	**use of grammar**
participants responding in English	3.10 (0.48)	2.85 (0.48)	3.25 (0.44)	2.80 (0.52)	3.40 (0.50)
participants responding in Chinese	3.00 (0.56)	2.65 (0.67)	3.15 (0.57)	2.85 (0.48)	3.10 (0.55)

In addition, 23 monolingual English and 16 monolingual Mandarin Chinese participants were recruited as control groups. They were between 18 and 25 years old. The English speakers were recruited at King’s College London, whereas the Chinese monolinguals were recruited at the China University of Geosciences, Beijing. Some of the participants indicated that they could speak another language, but when they were asked to evaluate their fluency and frequency of use, they reported this was at a basic level, being spoken occasionally, e.g., during holidays abroad. Hence, the decision was made to retain the data from those students for the final analysis, in particular, because it can be difficult to find ‘true’ monolingual speakers who are educated to a university level. In other words, for the purpose of this study, they were treated as ‘functional’ monolinguals.

Materials

To investigate the semantic domain of animals, 12 terms were selected, namely: *ant, cow, elephant, panda, camel, spider, bee, lion, monkey, butterfly, rabbit* and *tiger.* The items were arranged in such way that each word was compared with all others (A-B, A-C, A-D, (…), etc.), which resulted in a list of 66 pairs. Eight of the selected items were the same as those used by Romney et al. (1995) and Hermann and Raybeck (1981), whilst the remaining 4 items were chosen from the category of insects to avoid a clustering of the items. For the purpose of the Chinese version of the task, the selected terms were translated by a Chinese native speaker. The final list of the Chinese terms included: 蚂蚁 /mǎyǐ/, 母牛 /mǔniú/, 大象 /dàxiàng/, 熊猫 /xióngmāo/, 骆驼 /luòtuo/, 蜘蛛 /zhīzhū/, 蜜蜂 /mìfēng/, 狮子 /shīzi/, 猴子/hóuzi/, 蝴蝶 /húdié/, 兔子 /tùzi/, 老虎 /lǎohǔ/.

Design and procedure

The participants were asked to judge 12 animal terms, but instead of a four point scale as in the studies by Herrmann and Raybeck (1981) and Romney et al. (1995), we adopted a six point scale. This choice was motivated by two considerations: no inclusion of a mid-point and utilisation of a scale that is broad enough to offer a range of judgments. The task was presented to the participants in an electronic format (with the use of the LimeService platform) in such way that only one pair of terms (A-B) was visible at a time. The participants had to rank how similar (很相似 /hěn xiāngsì/) or dissimilar (很不相似 /hěn bù xiāngsì/) the pairs of animal terms are on a six point scale, with 1 referring to most dissimilar and 6 standing for most similar . The participants were not given any explicit instruction with regard to which dimension or characteristics they were supposed to evaluate the similarity of presented terms. Their attention was not directed to the size or habitat or any other predefined dimensions (e.g., shape, size, etc.). In this type of task, the set of dimensions that the participants rely on emerges on the map once the data have been analysed and the individual weights allow for evaluating the importance of the dimensions to each participant. The task, including the completion of a biographical questionnaire took, on average, 10 minutes. The participants were offered a box of chocolates or a Starbucks voucher (equal to 25HK$) for their time and effort.

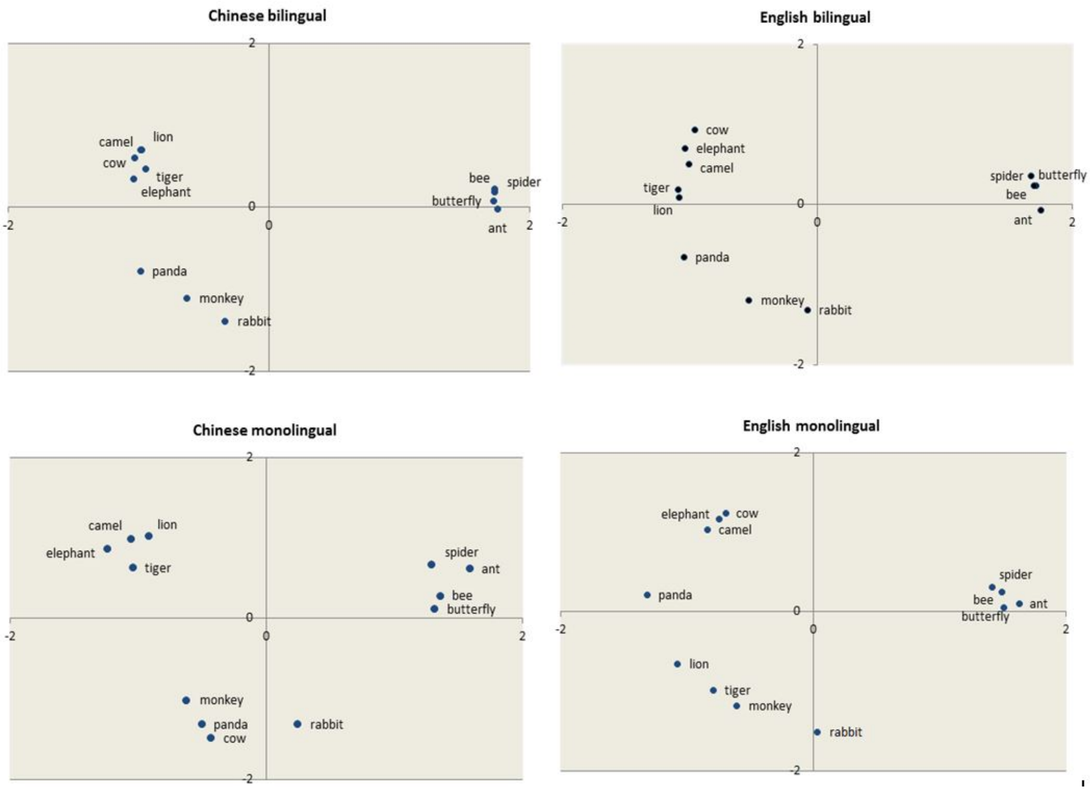
Results and discussion

First, the individual similarity judgments collected from each participant were entered into SPSS in the form of a 12x12 matrix having a lower triangular shape. The rows and columns were labeled with the animal terms. Also, a number 6 was entered on the diagonal where each of the terms met (e.g., ant-ant) to indicate the highest level of similarity. Next, the data were analysed using the ALSCAL MDS[1] algorithm (Takane, Young and DeLeeuw, 1977). Initially, four individual maps were produced, i.e., (1) a map of bilingual participants responding in Chinese, (2) a map of those bilinguals who responded in English, (3) a map of English monolingual, and (4) one of monolingual Chinese speakers (Figure 2). To produce each map, the matrices from each individual in each group of participants were stacked into a single matrix. Data from two bilingual participants responding in Chinese, five bilinguals providing answers in English, three English monolinguals and three Chinese participants, were removed from the final

[1] In SPSS the INDSCAL algorithm is included under the ALSCAL extension.

analysis as the participants either did not meet the selection criteria or provided unitary judgments for all pairs of words. This could have occurred due to a lack of understanding of the task or an unwillingness to attend to each stimulus separately. In sum, in the bilingual Chinese map, 18 matrices (from 18 participants) were stacked into a single one, thus resulting in 216 (18 x 12) rows and 12 columns. Subsequently, the averaged Kruskal stress value for this set of data was equal to .285; RSQ = .709. The same procedure was repeated for the bilingual English computations, but this time 15 matrices were stacked on top of each other, which resulted in one matrix of 15 x 12 = 180 rows and 12 columns, with the averaged stress value equal to .242: RSQ = .763. Furthermore, two separate monolingual maps were produced, i.e., an English one by stacking 20 individual matrices, which consisted of 20 x 12 = 240 rows and 12 columns (the averaged stress was calculated to be .344; RSQ = .404) and a Chinese one, produced by compiling 14 matrices together, which gave a matrix of 14 x 12 = 168 rows and 12 columns (the stress was estimated to be .398; RSQ = .202).

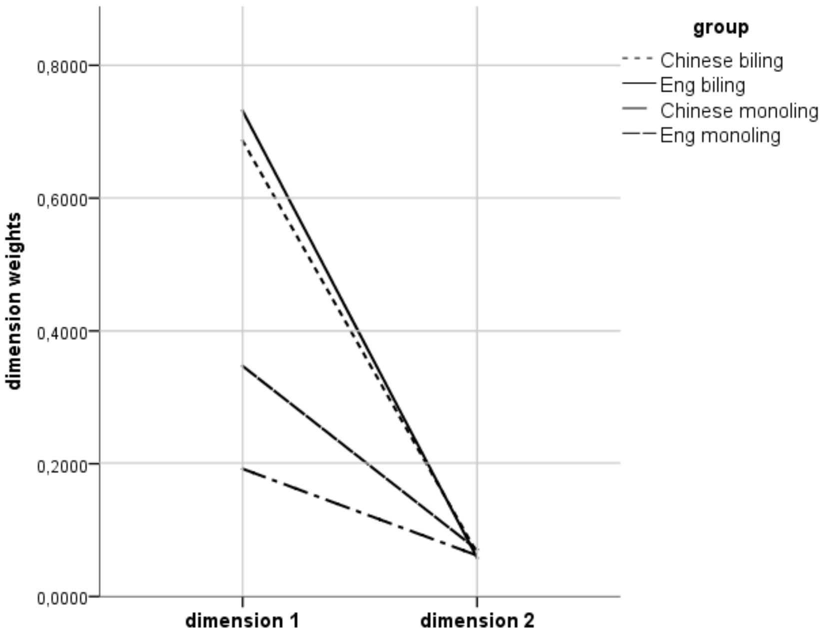
We have also calculated three and four dimensional solutions. The respective maps had the following Kruskal values: (1) bilinguals who provided responses in Chinese, 3D = .226; RSQ = .723, 4D = .179; RSQ = .726; (2) bilinguals who provided responses in English, 3D = .285; RSQ = .709, 4D = .168; RSQ = .717; (3) Chinese monolinguals, 3D = .265; RSQ = .230, 4D = .196; RSQ = .224; (4) English monolinguals, 3D = .252; RSQ = .427, 4D = .212; RSQ = .349. Despite the fact that the stress values for the 2D solutions were slightly higher than recommended (0.183 for 12 items following Sturrock and Rocha, 2000), we focused on the ease of interpretation of the maps following Bartholomew et al.'s (2002:63) suggestion that "there is a trade-off between improving fit [e.g., reducing Kruskal stress values] and reducing the interpretability of the solution". That is, three, four or more dimensional maps are increasingly more difficult to interpret and compare, especially taken a small number of items under consideration. Therefore, to retain the clarity of the presentations, 2D representations were preferred.



First of all, the visual dissimilarities were considered qualitatively. It can be seen that the Chinese and English bilingual maps show notable similarities. For example, in both languages the terms describing insects, i.e., *ant, bee, spider,* and *butterfly,* are clustered close together. Furthermore, the words *rabbit, monkey,* and *panda* are grouped together and the remaining five terms, *tiger, lion, camel, cow* and *elephant*, are in a close proximity on each map. However, we can also observe the proximities of the terms to the horizontal axis are reversed, i.e., on the English map, two terms, *tiger* and *lion*, are closer to the axis, whereas on the Chinese, these are further away from the axis, in turn, the terms referring to *cow* and *elephant* are closer. Also, it can be observed that there is a complete overlap between the two terms, *camel* and *lion*, in the Chinese version of the map. All in all, the animal terms presented on the Chinese and English bilingual maps have a similar distribution. A number of the terms are clustered to the extent that it is actually difficult to see the individual distributions.

The comparison of the semantic structure of animals of Chinese monolingual participants and bilingual speakers who responded in Chinese seems to reveal a greater degree of distribution than that reported for the previously described maps. That is, the distances between the individual terms are seemingly larger than those depicted on the Chinese and English bilingual maps. For instance, the term *cow* was judged by the bilingual and monolingual participants in a very different way. Finally, the comparison between the English monolingual participants and bilingual speakers who responded in English also suggests a certain level of distribution between the animal terms that is not, for instance, visible on the bilingual maps. The terms are located further away from each other, which could reflect the fact that the participants used a different set of judgments while evaluating the similarity of the given items. For example, differences are clearly visible between the terms: *panda*, *lion*, and *tiger*, which cross the horizontal axis.

Next, to investigate the dissimilarities quantitatively, we followed the analysis performed by Schrauf and Sanchez (2011) and used the INDSCAL (individual differences multidimensional scaling) algorithm to calculate weights that each participant gave to the two dimensions. This allowed for testing the differences in the two dimension weights between the four groups of participants (Chinese bilingual, English bilingual, Chinese monolingual, and English monolingual). The contrast between the two dimensions across the four groups is depicted in Figure 3.



Next a 4 x 2 mixed model analysis of variance (ANOVA) was conducted. The four groups (Chinese bilingual, English bilingual, Chinese monolingual and English monolingual) were treated as categories of the between subjects 'group' variable, and the two dimensions elicited from the semantic judgment task were the within subjects variable. The main effect of dimension was statistically significant, F (1, 66) = 136.73, $p < .001$. Overall, the participants paid greater emphasis to dimension one (M = .622, SD = .305) than dimension two (M = .226, SD = .117). Furthermore, a statistically significant main effect of group was observed, F (3, 66) = 9.45, $p < .001$ and a statistically significant two way interaction between the dimension and the group, F (3, 66) = 10.53, $p < .001$. The post hoc Tukey HSD test indicated that the bilingual and monolingual groups differed in their weights, i.e. the comparison of Chinese bilingual (M = .523, SD = .033) vs. Chinese monolingual judgments (M = .297, SD = .038) was statistically significant, $p < .001$, as well as the one between Chinese bilingual and English monolingual (M = .363, SD = .031) one, $p < .001$. The comparison of the English bilingual group (M = .490, SD = .033) to the other two monolingual ones also provided statistically significant differences, $p = .001$ (Chinese monoling) and $p < .05$

(English monolingual). Finally, no statistically significant difference was observed between the two bilingual groups, $p > .05$.

To sum up, the results presented above show that there is a level of similarity between the two bilingual maps that is significantly different from the English monolingual one and the Chinese monolingual map, as seen in the qualitative and quantitative analyses. Despite the fact that we employed a between subjects design, the items on the two bilingual maps are located in close proximities, which does not seem to be the case when the individual bilingual maps are compared to the monolingual-equivalents. This level of similarity between the two bilingual semantic structures might potentially point to a process of *semantic convergence*, i.e., a compound bilingual representation that draws on the information from both monolingual representations; a representation in which items become more alike (Ameel et al., 2005, 2009; Pavlenko, 1999). This is especially likely taken into account the relatively early age of English language acquisition among our bilingual participants.

Despite the fact that the findings of Study one are insightful, we acknowledge several limitations. First, a small number of prototypical terms was used in the task, i.e., only 12 terms. Furthermore, animal terms are not represented on a continuum, unlike, for instance, household objects or containers that were investigated by e.g., Ameel et al. (2005; 2009). Therefore, the way in which speakers of different languages judge the similarities between various animal terms might have been too small to detect. Finally, the between subject design may render some issues with regard to the comparison of the MDS maps. In order to address these limitations and to extend the findings to other semantic domains, a second Study investigating emotion terms with a within subject as well as between subject design was conducted.

Study two

Method

Participants

Data collection took place at the same sites as for Study one. Also, all bilingual participants were enrolled on undergraduate and postgraduate courses, were between 18 and 30 years old and were born in Mainland China. The English language proficiency score was collected in the form of self-ratings of the five main language skills (Table 2). The age range of English acquisition for this group of speakers varied from 3 years of age to 14 ($M = 7.6$). In total, a group of 16 bilingual speakers was requested to perform the similarity judgment on two separate occasions, once in Chinese and once in English or vice versa, about one week apart.

	listening	**speaking**	**reading**	**writing**	**use of grammar**
English	3.35 (0.587)	3 (0.562)	3.45 (0.511)	3.1 (0.553)	3.35 (0.587)

In addition, two groups of monolingual English and Chinese speakers were recruited to act as a control groups. In the UK, data from 26 participants were collected; however, two data sets were not used in the final analysis due to the fact that the participants indicated being proficient in Hebrew and Gujarati, hence not qualifying as monolingual. Also, 21 students between the ages of 18 and 25 provided responses to the Chinese task; however, data from only 17 was retained for the final analysis as four participants indicated being fluent in English. Once again, we refer to both monolingual groups as functional monolinguals, because they indicated having some knowledge of other languages (e.g., English, Korean, Japanese, German, French, Mongolian, Russian, Spanish and Uighur regarding the Chinese participants and for English speakers the selection included: French, German, Korean, Spanish and Portuguese), which they used occasionally on holidays or with friends. All participants received an e-voucher or cash as a form of gratification for taking part in the study. The participants in Hong Kong received a Starbucks voucher (45HK$); whereas, participants in Beijing and London were given an Amazon voucher or cash equivalent of 4 Euros.

Materials

A list of 20 emotion terms was initially selected in English, including: *anger, anxiety, boredom, disgust, envy, sadness, shame, annoyance, worry, disappointment, excitement, happiness, love, hope, pride, courage, trust, calmness, satisfaction, and relief*. The terms were selected in such way that 10 represented positive and 10 negative emotions. In addition to the 'valence' or 'intensity' of each emotion was taken into account. The chosen emotions represented: negative and forceful (anger, disgust, annoyance), negative and not in control (anxiety, worry), negative and passive (boredom, sadness, shame, disappointment), negative thoughts (envy), positive and lively (excitement, happiness), caring (love), positive thoughts (hope, pride, courage, trust, satisfaction), quite positive (calmness, relief) forms. This selection was performed according to the Emotion Annotation and Representation Language (EARL) proposed by the Human-Machine Interaction Network on Emotion (HUMAINE). Such classification of emotion terms was included as part of the preparation stage in order to enable easier interpretation of the maps. Next, the list of all terms was translated into Chinese by a native Chinese speaker. The Chinese items included: 愤怒/fènnù/, 焦虑/jiāolǜ/, 厌倦/yànjuàn/, 厌恶/yànwù/, 嫉妒/jídù/, 悲伤/bēishāng/, 羞耻/xiūchǐ/, 烦恼/fánnǎo/, 担心/dānxīn/, 失望/shīwàng/, 兴奋/xīngfèn/, 幸福/xìngfú/, 爱/ài/, 希望/xīwàng/, 自豪/zìháo/, 勇气/yǒngqì/, 信赖/xìnlài/, 宁静/níngjìng/, 满足/mǎnzú/, 安慰/ānwèi/. Finally, two versions of the task were prepared, one in English and one in Chinese, in which the emotion terms were arranged into pairs in such a way that each was compared with all the others. This resulted in the comparison of 190 pairs of words; presentation of which was randomized for every participant by the online tool (the LimeSurvey).

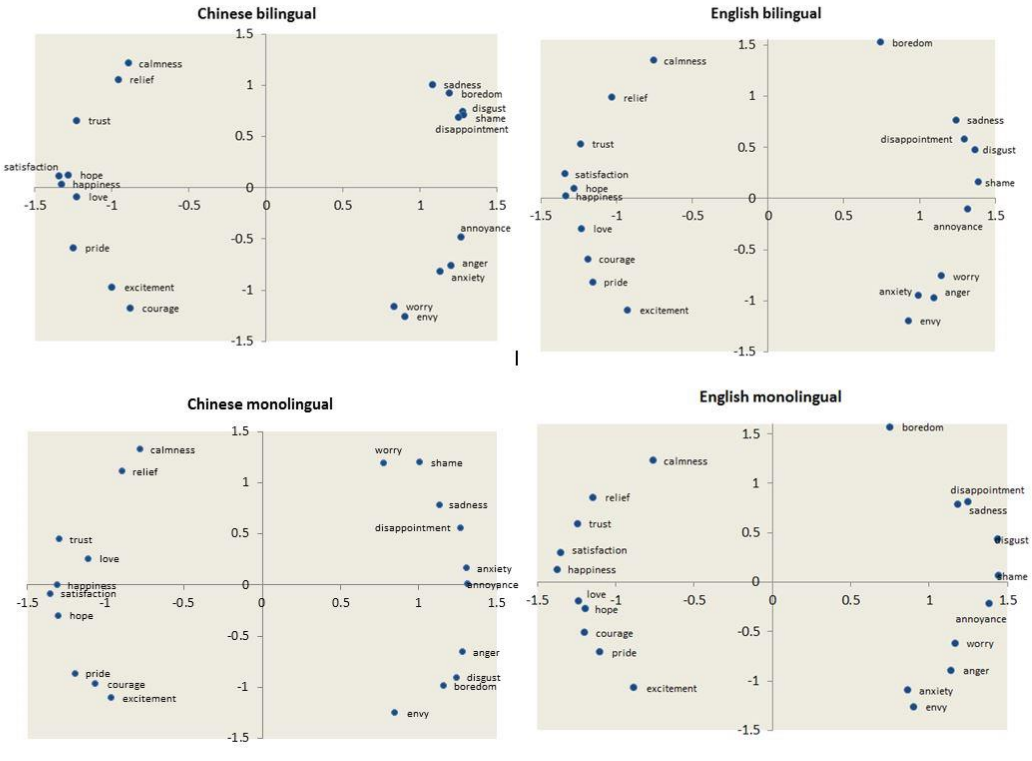
Design and procedure

The procedure employed in this Study was identical to that used in Study one.

Results and discussion

Data preparation followed the same procedure as employed in Study one. All maps were analysed to start with using the ALSCAL MDS algorithm (Takane, Young and DeLeeuw, 1977) in SPSS. Four individual maps were created, i.e., (1) bilingual judgments given in English; (2) bilingual judgments provided in Chinese; (3)

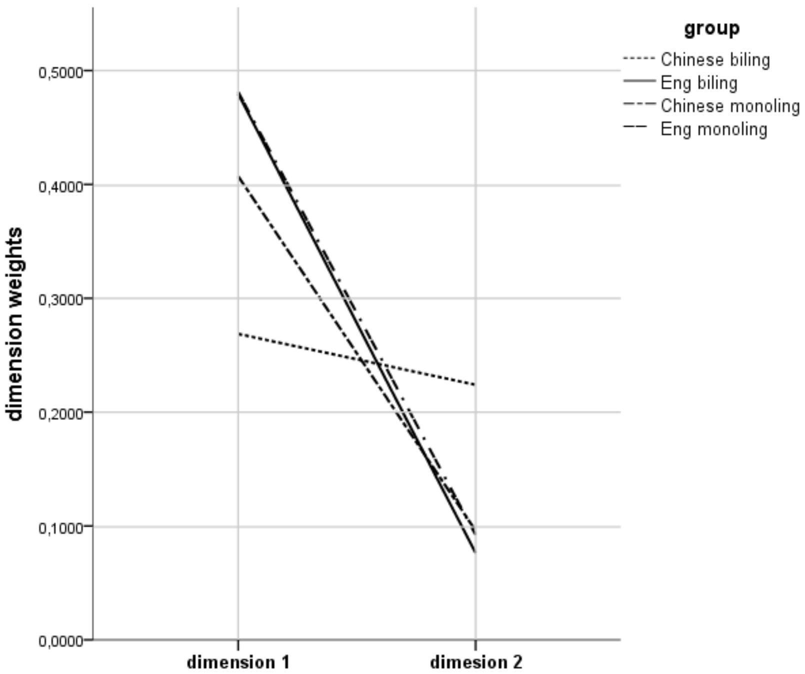
monolingual English and (4) monolingual Chinese ones (Figure 4). Both bilingual maps were created based on stacking 16 matrices into two single ones, resulting in a 320 rows x 20 column matrix with the Kruskal stress values being .327; RSQ = .472 for the Chinese map and .306; RSQ = .530 for the English one. The map, which was based on the Chinese monolingual data, was created from 17 individual matrices that were stacked (340 rows x 20 columns). The stress value for this map was .308; RSQ = .499. Finally, the English monolingual map was based on 24 matrices (480 rows x 20 columns) with a stress value of .295; RSQ = .558. Sturrock and Rocha (2000) advise a stress value of .279 for 20 items in 2D solutions. Therefore, in addition 3D and 4D solutions were computed and the reported Kruskal stress values were: (1) bilingual judgments given in English, 3D = .243; RSQ = .512, 4D = .199; RSQ = .487; (2) bilingual judgments provided in Chinese, 3D = .258; RSQ = .584, 4D = .206; RSQ = .656; (3) monolingual English, 3D = .229; RSQ = .235, 4D = .186; RSQ = .427; (4) monolingual Chinese, 3D = .23882; RSQ = .513, 4D = .209; RSQ = .502. In spite of lower stress values and higher RSQ statistics, which points to the amount of variance accounted for by the number of dimensions, for the 4D maps, for ease of interpretation we focus here on the 2D maps.



The qualitative investigation of the distribution of terms on the maps (Figure 4) clearly illustrates a division between the negative emotions to the right and the positive ones to the left of the y axis. Consequently, the x axis could be identified as representing ‘valence’. The y axis cannot, unfortunately, be identified unequivocally as depicting ‘forcefulness’ or ‘intensity’ of emotions. Some of the words do cluster according to the a priori specified categories. For instance, *calmness* and *relief* are close to one another, representing the category of positive and quiet emotions, but other positive emotion terms are outside of the prescribed categories. Regarding which, the participants judged *excitement* and *pride* as being very similar to one another. A similar picture can be observed with regard to the negative emotions. For, *anger* and *annoyance* are both negative and forceful, however, *disgust*, which also belongs to the same category, is further away.

The comparison of the bilingual responses provided in Chinese and English reveals a degree of similarity, i.e., terms referring to the same emotion are close to each other e.g., *calmness* and *relief*. Also, terms referring to *hope* and *happiness* are almost fully overlapping. An even greater degree of similarity can be observed on the map comparing English bilingual responses with those of English monolingual ones. The distances between individual data points appear to be very short, which could indicate that bilingual speakers have internalised English patterns of judging emotion terms. The greatest dissimilarity of terms can be observed on the maps of the Chinese bilingual group and the Chinese monolingual one. A number of terms are located in opposing cells, e.g., *disappointment, anxiety, boredom* and *shame*. Furthermore, the distances on those maps are seemingly larger. This could be indicative of the greater differences in judgments of emotion terms between these two groups of participants.

The next stage of the analysis involved exploring the differences in the two dimension weights that participants gave during the semantic judgement tasks (Figure 5). The main effect of dimension was statistically significant, F (1, 69) = 369.26, $p < .001$. Greater weights were attributed to the dimension one (M = .618, SD = .018) rather than dimension two (M = .332, SD = .009). However, no main effect of group was observed, F (3, 69) = .293, $p < .001$. The interaction between the main two factors, on the other hand, turned out to be statistically significant, F (3, 69) = 29.328, $p < .001$. Nevertheless, a post hoc Tukey HSD comparison did not reveal any statistically significant differences in weights between the four groups. In all individual comparisons the p value was above .05.



In sum, the emotions on all four maps are arranged in a circular order. However, we observed small differences with regard to the distribution of terms on the maps in terms of placement as well as distance. The greatest level of similarity was observed between the English bilingual and the English monolingual maps. The bilingual maps revealed a fair amount of similarity. Similarly to the findings obtained in the Study one, the greatest dissimilarities were observed between the Chinese monolingual map and the remaining ones. The above findings not only may point to a process of semantic convergence as, but also to a possible 'shift away' from the L1 semantic structure towards English-native-like one. This finding can potentially be related to the level of English language proficiency of this group of participants. All of the above findings are further discussed in the forthcoming section.

Discussion

The semantic structure of animal terms and emotion terms was investigated with groups of bilingual Mandarin Chinese-English speakers and monolingual Chinese and English speakers. The investigation of the semantic domain of animals in Study one has revealed that the items on the two bilingual maps were seen by the participants as being very similar to each other, which might be indicative of the process of semantic convergence; a representation in which the items from L1 and L2 converge to a common pattern. This process can be driven by early language acquisition, as previous studies have shown (e.g., Ameel et al., 2005; or Kersten et al., 2010). Furthermore, the distribution of the animal terms on the bilingual maps has appeared not to be the same as the monolingual ones, which suggests that the bilingual participants viewed the items differently from the monolingual speakers. The largest degree of dissimilarity was recorded between the Chinese bilingual map and the Chinese monolingual one, which could indicate that the bilingual representation undergoes a change; it 'shifts away' from the monolingual L1 representation, possibly due to the influence of L2. The findings obtained in Study two, in which the semantic structure of emotion terms was investigated, have pointed towards a similar pattern of results to those observed in Study one. That is, it has been demonstrated that there is a level of similarity between the two bilingual maps. The greatest dissimilarity has once again been reported between the Chinese bilingual map and the Chinese monolingual one. However, the shortest distance, this time, was seen between the English bilingual and the English monolingual maps. Finally, the emotion terms, on all maps, fell in a circular order, and one of the dimensions can be interpreted as positive-negative. This set of findings is in line with the highly influential study conducted by Russell (1983).

Based on the combined findings from both Studies we can make several general observations. First of all, the bilingual maps have revealed that the bilingual semantic structures of animal terms and emotion terms are dynamic and they seem to undergo the process of semantic convergence (e.g., Ameel et al., 2005, 2009; Pavlenko, 1999). Furthermore, in both studies the largest dissimilarity has been observed between the Chinese bilingual and the Chinese monolingual maps. It appears that the Chinese bilingual semantic structures are subject to a change due to the impact of L2. This can be seen as evidence of a semantic shift towards L2. Also, it appears that the English bilingual representation approximates to that of English monolingual one. Finally, we cannot confirm the theoretical proposal made by Gentner and Boroditsky (2001) regarding the cross-linguistic variation between abstract and concrete domains. We have not observed a greater variation between the emotion terms as compared to the animal terms. This, however, might have to do with the selection of prototypical exemplars.

In sum, the findings obtained in those studies can be interpreted as pointing towards a process of bilingual semantic change in the form of convergence and semantic shift towards L2. Pavlenko (1999:223) explained that the process of semantic convergence refers to when “a unitary system is created, [which is] distinct both from L1 and L2”. The convergence may take place due to a parallel activation of either languages, or more specifically, due to a process of *retrieval-induced reconsolidation* (Wolff and Ventura, 2009). That is, it is likely that one memory trace becomes labile and susceptible to change, for a limited period of time, once it has been reactivated by another trace (Ameel et al., 2009:272). Since both bilingual groups of participants in the current study were proficient users of Chinese and English and because they used both languages on a daily basis, it is possible that “the encounters in each language may reactivate the other language frequently, resulting in labile memory traces that are susceptible to cross-linguistic interference in both directions”, as put forward by Ameel et al. (2009:272). In turn, this long-term interaction between L1 and L2 can lead to a somewhat intermediate system of representation, in the form of a semantic interlanguage. Furthermore, the fact that the bilingual groups of participants who were proficient L2 users, living in a mixed language environment (in Hong Kong both English and Mandarin Chinese are routinely used) and using English for academic purposes, could have resulted in a ‘shift away’ of the Chinese semantic structure from the monolingual Chinese one. This once again points to 1) the fact that the bilingual semantic structures can be seen as dynamic and 2) the uniqueness of bilingual speakers.

Our proposal regarding the existence of semantic interlanguage is still underspecified given the fact that the current set of studies is the first direct attempt of investigating this notion and given the fact that our results are not uniform. We envisage the semantic interlanguage to integrate elements of L1 and L2 semantic systems. It is also likely that one of the systems, say L2, is to a large extent embedded in the L1 representation, especially if the first language is the more dominant one or if the languages were acquired in a successive manner. However, other factors such as: age of acquisition, language proficiency, frequency of language use, length of exposure or age of active language use may lead to differing semantic structures, and hence should be incorporated into future investigations. Currently, we are in the process of conducting a study on the semantic domain of emotions with bilingual German-English speakers with different levels of L2 proficiency. Our working hypothesis is that the semantic patterns of the more proficient speakers should be more like the monolingual English ones when compared to those obtained from less proficient speakers. Finally, it has to be acknowledged that both studies utilised a small number of prototypical terms. This potential shortcoming is imposed by the pair-wise methodology. That is, if 20 terms are arranged in pairs in such a way that each item is compared with every other, this results in 190 pairs. Having a large number of items to evaluate on a scale may lead to tiredness and boredom; this could potentially result in participants not attending to them in the required manner. This issue could have caused the poor goodness of fit of our data (high Kruskal stress values for 2D maps). Therefore, in future investigations, the number of items should be increased focusing not only on prototypical exemplar but also those peripheral ones, if possible, represented on a continuum. Finally, for investigations employing larger stimuli list the sorting procedure used e.g., by Holmes and Wolff (2013) might be preferred.

Conclusions

The following studies aimed to illustrate the process of a bilingual semantic change through the investigation of the semantic structure of animal and emotion terms. It has been demonstrated that the bilingual L1 and L2 semantic representations appear very similar and could potentially exemplify a process of semantic convergence. Furthermore, it has been shown that bilingual Chinese representations differ to a reasonable extent from the Chinese monolingual ones. This might be illustrative of a semantic shift towards L2. However, since the notion of semantic interlanguage is very new, it yet has to be verified. Nonetheless, the notion of interlanguage is an important one as it points to the fact that learning a second, third, or a fourth language should not be seen as striving for native-like performance, but rather, as learning a new way of labeling, referencing and categorising the world along with learning a new set of values and cultural norms that results in a bilingual individual becoming a multi-competent L2, L3, or L4 user (Cook, 2003) who does not fully resemble either of the monolingual speakers.

References

Alvarado, N., and Jameson, K. A. (2011). Shared Knowledge About Emotion Among Vietnamese and English Bilingual and Monolingual Speakers. *Journal of Cross-Cultural Psychology, 42*(6), 963-982.

Ameel, E., Malt, B. C., Storms, G., and Van Assche, F. (2009). Semantic convergence in the bilingual lexicon. *Journal of Memory and Language, 60*(2), 270-290.

Ameel, E., Storms, G., Malt, B. C., and Sloman, S. A. (2005). How bilinguals solve the naming problem. *Journal of Memory and Language, 53*(1), 60-80.

Bartholomew, D. J., Steele, F., Moustaki, I., and Galbraith, J. (2002). *The analysis and interpretation of multivariate data for social scientists*: Chapman and Hall Ltd.

Cook, V. (2003). *Effects of the Second Language on the First* (Vol. 3). Clevedon [u.a.]: Multilingual Matters.

Fillenbaum, S., and Rapoport, A. (1971) *Structures in the subjective lexicon*. New York: Academic Press.

Dickerson, L. J. (1975). The learner's interlanguage as a system of variable rules. *TESOL quarterly*, *9*(4), 401-407.

Gass, S. (1984). A review of interlanguage syntax: Language transfer and language universals. *Language learning, 34*(2), 115-132.

Gentner, D., and Boroditsky, L. (2001). Individuation, relativity, and early word learning. In M. Bowerman and S.C. Levinson (Eds.), *Language acquisition and conceptual development*, (pp. 215-256). Cambridge, UK: Cambridge University Press.

Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition, 1*(2), 67-81.

Grosjean, F. (1989). Neurolinguists, beware! The bilingual is not two monolinguals in one person. *Brain and Language, 36*(1), 3-15.

Harrison, G. L., and Krol, L. (2007). Relationship between L1 and L2 word-level reading and phonological processing in adults learning English as a second language. *Journal of Research in Reading, 30*(4), 379-393.

Holmes, K. J., and Wolff, P. (2013). Spatial language as a window on representations of three-dimensional space. *Behavioral and Brain Sciences, 36*(5), 550-551.

Herrmann, D. J., and Raybeck, D. (1981). Similarities and differences in meaning in six cultures. *Journal of Cross-Cultural Psychology, 12*(2), 194-206.

Jared, D., Poh, P. Y., and Paivio, A. (2013). L1 and L2 picture naming in Mandarin–English bilinguals: A test of Bilingual Dual Coding Theory. *Bilingualism: Language and Cognition, 16*(2), 383-396.

Kasper, G., and Schmidt, R. (1996). Developmental issues in interlanguage pragmatics. *Studies in Second Language Acquisition, 18*(02), 149-169.

Kersten, A. W., Meissner, C. A., Lechuga, J., Schwartz, B. L., Albrechtsen, J. S., and Iglesias, A. (2010). English speakers attend more strongly than Spanish speakers to manner of motion when classifying novel objects and events. *Journal of Experimental Psychology: General*, *139*(4), 638.

MacWhinney, B. (1997). Second language acquisition and the competition model. In A. M. B. de Groot and J. F. Kroll (Eds.), *Tutorials in bilingualism: Psycholinguistic perspectives* (pp. 113-142). Mahwah: Lawrence Erlbaum Associates.

Moore, C. C., Romney, A. K., Hsia, T., and Rusch, C. D. (1999). The universality of the semantic structure of emotion terms: Methods for the study of inter and intra cultural variability. *American Anthropologist, 101*(3), 529-546.

Pavlenko, A. (1999). New approaches to concepts in bilingual memory. *Bilingualism: Language and Cognition, 2*(3), 209-230.

Pavlenko, A. (2009). *The bilingual mental lexicon: interdisciplinary approaches*. Bristol: Multilingual Matters.

Romney, A. K., Batchelder, W. H., and Brazill, T. (1995). Scaling semantic domains. In D. R. Luce, M. D'Zmura, D. Hoffman, G. J. Iverson and A. K. Romney (Eds.), *Geometric representations of perceptual phenomena: Papers in honor of Tarow Indow's 70th birthday* (pp. 267-294). Hillsdale: Lawrence Erlbaum.

Romney, A. K., and Moore, C. C. (1998). Toward a theory of culture as shared cognitive structures. *Ethos, 26*(3), 314-337.

Romney, A. K., Moore, C. C., and Rusch, C. D. (1997). Cultural universals: Measuring the semantic structure of emotion terms in English and Japanese. *Proceedings of the National Academy of Sciences of the United States of America, 94*(10), 5489-5494.

Russell, J. A. (1983). Pancultural aspects of the human conceptual organization of emotions. *Journal of Personality and Social Psychology*, *45*(6), 1281.

Schrauf, R. W., & Sanchez, J. (2011). The shifting structure of emotion semantics across immigrant generations: Effects of the second culture on the first language. *Modeling bilingualism: From structure to chaos*, 177-200.

Selinker, L. (1972). Interlanguage. *IRAL-International Review of Applied Linguistics in Language Teaching*, *10*(1-4), 209-232.

Swan, M. (1997). The influence of the mother tongue on second language vocabulary acquisition and use. *Vocabulary: Description, acquisition and pedagogy*, 156-180.

Sturrock, K., & Rocha, J. (2000). A multidimensional scaling stress evaluation table. *Field methods*, *12*(1), 49-60.

Takane, Y., Young, F. W., & De Leeuw, J. (1977). Nonmetric individual differences multidimensional scaling: An alternating least squares method with optimal scaling features. *Psychometrika*, *42*(1), 7-67.

Tsai, J. L., Simeonova, D. I., and Watanabe, J. T. (2004). Somatic and social: Chinese Americans talk about emotion. *Personality and social psychology bulletin*, *30*(9), 1226-1238.

Wierzbicka, A. (1992). Talking about emotions: Semantics, culture, and cognition. *Cognition and Emotion, 6*(3-4), 285-319.

Wolff, P., and Ventura, T. (2009). When Russians learn English: How the semantics of causation may change. *Bilingualism: Language and Cognition*, *12*(2), 153-176.

Yoshida, M., Kinase, R., Kurokawa, J., and Yashiro, S. (1970). Multidimensional scaling of emotion. *Japanese Psychological Research*, *12*, 45-61.